Full Factorial Design: Identifying Which Players Were Built for the NBA Bubble

Group 2

1. Introduction & Motivation

The COVID-19 pandemic brought with it many challenges in the sports world. Sports have always revolved around the fans, but unfortunately, events involving fans often result in “super-spreader” events that worsen the pandemic. The National Basketball Association (NBA) was one of many sport leagues that had their season interrupted by the pandemic. Their solution was to implement a contained, quarantined environment in Orlando, Florida called the “Bubble”, where game-play could continue under strict precautionary procedures.

In this investigation, we were interested in comparing NBA player performance inside and outside the Bubble. The Bubble provided a contrast to the usual regime of home and away games. Since all teams competed on the same court without fan presence, it is conceivable that the effect of home and away games on player performance might have changed in the Bubble. For a few players, this perceived difference between their normal play and their Bubble play seemed especially pronounced. These differences were seized upon by the media and developed into influential narratives that shaped the 2020 NBA postseason. As such, we were particularly interested in assessing whether the Bubble effect was real for these specific players, and whether these effects conform to the media narrative.

The primary objective of our investigation was to identify the effects of home and Bubble games on player performance. The Bubble provided an unprecedented opportunity to examine the impact of fans and packed arenas on professional sport outcomes by providing a controlled, identical setting for competition. Using a full factorial design, we tested the significance and magnitude of playing at home, playing in the Bubble, and their interaction.

2. Methods

2.1. Data

To begin our data analysis, we selected six different NBA players to analyze based on their widespread media coverage. Three of the players had acquired a reputation as “underperformers” in the Bubble, while the three others had supposedly “overperformed” or surpassed expectations in the Bubble. Our underperformers were Paul George, Pascal Siakam, and Giannis Antetokounmpo, and our overperformers were Jamal Murray, Donovan Mitchell, and Jimmy Butler. All data was collected from Basketball Reference, a popular online database of basketball statistics from professional basketball players around the world. Per-game advanced statistics for the 2019-2020 season (which included pre-Bubble and in-Bubble information) were downloaded for each player.

Excel was then utilized to format the dataset for input to JMP Pro 15, the software we used for our statistical analysis. Empty values were replaced and the independent factors *home* and *bubble* were converted to binary variables with values 0 or 1. Finally, the six separate datasets were ready to be analyzed in JMP. The dataset can be found in the Github link provided in the appendix.

2.2. Model & Variables

Our two independent factors were home and bubble. The three responses of interest were offensive rating (ORTG), defensive rating (DRTG), and usage rate (USG%). Their models are shown as follows, respectively:

$$y_{ijk}^{Ortg} = u_{ij}^{Ortg} + \varepsilon_{ijk}^{Ortg}, \quad \text{where} \quad u_{ij}^{Ortg} = u^{Ortg} + \alpha_i^{Ortg} + \beta_j^{Ortg} + \alpha\beta_{ij}^{Ortg}$$

$$y_{ijk}^{Drtg} = u_{ij}^{Drtg} + \varepsilon_{ijk}^{Drtg}, \quad \text{where} \quad u_{ij}^{Drtg} = u^{Drtg} + \alpha_i^{Drtg} + \beta_j^{Drtg} + \alpha\beta_{ij}^{Drtg}$$

$$y_{ijk}^{Usg} = u_{ij}^{Usg} + \varepsilon_{ijk}^{Usg}, \quad \text{where} \quad u_{ij}^{Usg} = u^{Usg} + \alpha_i^{Usg} + \beta_j^{Usg} + \alpha\beta_{ij}^{Usg}$$

$$\text{where} \quad i = 1,2 \quad j = 1,2 \quad k = 1,...,n_{ij}$$

y_{ijk}^{Ortg} represents the response offensive rating, which refers to the estimate of points produced or scored per 100 possessions. y_{ijk}^{Drtg} is the defensive rating which is the estimate of points allowed per 100 possessions. Finally, y_{ijk}^{Usg} represents the usage rate, which refers to the estimate of the percentage of team plays that used the player while he was on the court. A higher offensive rating, a usage rate closer to 100%, and a lower defensive rating all correspond with better performance.

u is the overall mean. α_i is the effect of the away/home game; in our dataset the two levels are 0 and 1, corresponding to away and home, respectively. β_j is the Bubble effect; in our dataset the two levels are 0 and 1, corresponding to out-of-Bubble and in-Bubble, respectively. $\alpha\beta_{ij}$ is the interaction effect between home and Bubble. The model has a random error term $\varepsilon \sim N(0, \sigma_\epsilon^2)$ with all ε independent. For testing significance for each effect, our null hypothesis was that all of the effects are equal to zero, and our alternative hypothesis was that at least one of the effects was not equal to zero. All tests were conducted at the $\alpha = 0.05$ level.

2.3. Model Justification & Assumptions

A full factorial model was chosen because it can reveal interactions and is more efficient than testing one factor at one time. Since the levels of our factors are discrete, they lend themselves naturally to a 2^2 factorial design. A full factorial model was fit for each player using the "Fit Model" feature in JMP. Interaction plots for each player-response were generated to visually inspect the interactions (if present) between the effect of home and Bubble games. ANOVA tables were used to assess model significance and effect sizes and significance were estimated using Least Squares Mean Contrast.

A full factorial model rests on two assumptions: normality of residuals and equal variances across treatment combinations. Using JMP again, we checked these assumptions. For normality of residuals, we conducted a Shapiro-Wilk test for each player-response combination. Both of these tests were conducted at a significance level of 0.05. All of our JMP scripts can be found in the Github link provided in the appendix.

3. Results

In this section, we discuss the *significant* results from our analyses. Overall, most significant effects were from the Bubble; no players had any significant home effects and only one player had a significant interaction effect. In addition to the JMP figures given in this section, a complete table of effect estimates and p-values for all six players across all effects and responses is given in the appendix. All players met normality and equal variances assumptions except for Paul George, whose usage rate residuals were not normally distributed, with p-value<0.0039.

3.1. Giannis Antetokounmpo

Giannis demonstrated increased defensive rating in the Bubble; the effect was significant at the 95% confidence level, with a p-value of 0.0151. On average, Giannis' defensive rating increased by 8.05

(with a standard error of 3.2317) from no-Bubble to in-Bubble, indicating that Giannis performed worse at the defensive end while in the Bubble.

For usage rating, the interaction effect between home and Bubble was significant, with a standard error of 1.8246 and a p-value of 0.0414. The interaction plot in *Figure 1* shows that Giannis' usage rate increases in pre-COVID arenas when going from an away game to a home game. However, Giannis' usage rate decreases in the Bubble comparing away games to home games.

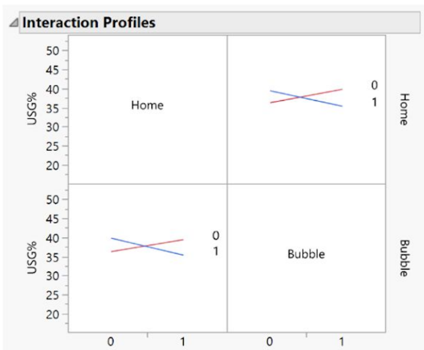

Figure 1: Interaction Plot of USG% for Giannis Antetokounmpo

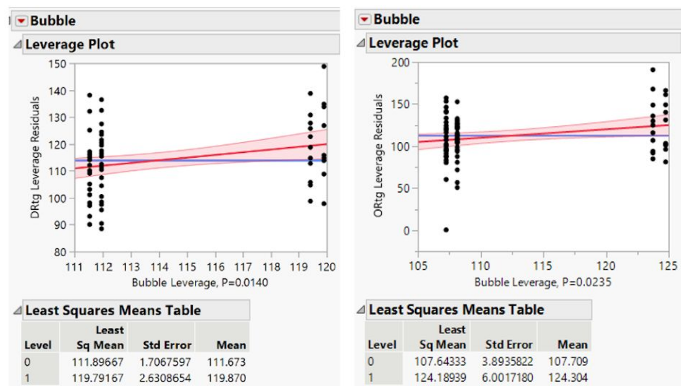

Figure 2: Leverage Plot and LS Means Table for Jamal Murray's (left) Drtg in the Bubble and (right) Ortg in the Bubble

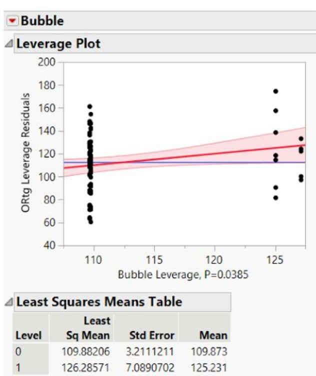

Figure 3: Leverage Plot and LS Means Table for Donovan Mitchell's ORtg in the Bubble

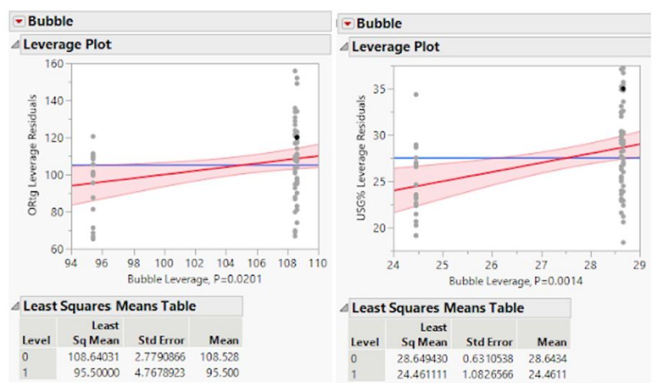

Figure 4: Leverage Plot and LS Means Table for Pascal Siakam's (left) ORtg and (right) USG% in the Bubble.

3.2. Jamal Murray

Jamal Murray had a significant Bubble effect for offensive rating, with an effect estimate of 16.546 and a standard error of 7.1541, as seen in *Figure 2*. He also had a significant Bubble effect for defensive rating, with an effect estimate of 7.895 and a standard error of 3.136.

3.3. Donovan Mitchell

Donovan Mitchell had a significant Bubble effect for offensive rating, having a p-value of 0.0385. The effect estimate was 16.404 with a standard error of 7.7824, as shown in *Figure 3*.

3.4. Pascal Siakam

Figure 4 shows that Pascal Siakam had a significant Bubble effect for both offensive rating as well as usage rate, having a p-value of 0.0201 and 0.0014, respectively. The effect estimate for offensive rating was -13.14 with a standard error of 5.5187. For usage rate, the effect estimate was -4.188 with a standard error of 1.2531.

4. Discussion

4.1. Giannis Antetokounmpo

Interestingly, Giannis was more of a defensive presence during pre-COVID arena environments compared to in the Bubble. Giannis was awarded the prestigious Defensive Player of the Year (DPOY) award for this past 2019-2020 season. Since this honor is given based on performance in the regular season (which corresponds to pre-Bubble conditions), it seems that perhaps without the motivating energy of live crowds, Giannis, along with the rest of his team, underperformed.

Furthermore, in normal crowd-packed arenas, Giannis exhibited a higher usage rate at home than away, whereas in the Bubble, this trend was reversed. This suggests that Giannis' play is indeed highly influenced by the presence of fans. Another possibility is that, as a home-crowd favorite, Giannis has a tendency to receive more usage in the presence of home fans. Either way, this interaction strongly indicates that Giannis is a player whose performance was indeed negatively affected by the Bubble.

4.2. Donovan Mitchell

Mitchell was rightfully classified as an over-performing player. Mitchell's exceptional offensive improvement seemed to have been adequately captured by the media narrative. It is worth mentioning that there was an outlier with a value of 199 in Mitchell's offensive rating. With this data point removed, Mitchell's offensive rating effect was no longer significant. However, since this data point arises from what was hailed by the media as a legitimately extraordinary performance, and not data-entry error or statistical misrepresentation, we did not remove it.

4.3. Jamal Murray

Prior to the Bubble, Jamal Murray was known to be a very unpredictable, albeit talented, player. In the Bubble, however, he had a number of exceptional performances. Commensurately, media reporting was positive, and he was widely touted as one of the top over-achievers of the Bubble. The statistical analysis, in this case, partially agrees with the media's conclusion. While Murray did experience significant improvements in *offensive* rating from pre-Bubble conditions, there were apparent deficiencies in his performance on the *defensive* end. The net effect, based on effect estimates, seems to have been strongly positive, but it is possible that the media overrated his impact. It is not uncommon for offensive production to be emphasized and valued over defensive production. Thus a player like Murray who experiences significant improvements on the offensive end concurrent with significant, but lesser magnitude declines on the defensive end, is likely to earn favorable media coverage.

4.4. Paul George

On the other hand, one of the most reviled players of the Bubble was Paul George. Regarded as a superstar, his failure to deliver while playing in the Bubble resulted in extremely negative media coverage. However, based on the lack of significance of any effects in our model, it is apparent that his poor Bubble performance was overstated. For the responses that we chose, it is apparent that neither Bubble nor home game status had any effect on George's gameplay. This, of course, does not rule out true underperformance, since there are other ways to measure performance besides the three responses we modeled.

4.5. Pascal Siakam

Pascal, one of the previously-identified underperformers, showed results that backed up what the media was saying about his sub-par performance. His significant decrease in both offensive rating and usage rate in the Bubble is indicative that the moniker that was originally assigned to Paul George: "Pandemic P", should actually have been given to Pascal Siakam. Siakam's team, the defending champion Toronto Raptors, also experienced underwhelming early playoff elimination. The lackluster offensive rating of Siakam and his subsequent lack of impact on his team's success (as seen in his decreased usage rate) is exacerbated by the fact that he was the league's Most Improved Player just one season prior to the onset of the Bubble. We see from our analysis that media criticisms of Pascal were likely warranted.

5. Final Thoughts

It is crucial to note that there are a number of potential confounding factors that must be acknowledged when interpreting our results. For example, the majority of Bubble games for players were playoff games. In our model, "playoff" could not be included as its own factor because there were no observations for playoff games that were not in the Bubble. Another fact to note is that our response variables are not comprehensive. While significant differences in these metrics should indicate a significant effect for our factor variables, the absence of significant differences for these metrics does not imply that the player's performance in general was not affected by the Bubble.

Regardless, we have shown that the media's perception of NBA player performance is overall accurate. The Bubble environment seems to have had a statistically significant effect on certain players, and the nature of the impact, negative or positive, certainly influenced the performance of that player's team in their playoff run.

With the coronavirus pandemic affecting the entire world, many venues, processes, and organizations have switched to "new normals" which try to provide some measure of normalcy under otherwise unprecedented circumstances. Despite this, our analyses of NBA athletes show that the "new normal" does indeed impact certain people differently, which is, more broadly, an issue that all employers should keep in mind while constructing safe environments for employees during this pandemic.

6. Appendix

6.1. Link to Data and JMP Scripts

https://github.com/isaacke9/built-for-the-bubble

6.2. Table of Effect Estimates and P-Values

	ORTG - Effect size	ORTG - P-value	DRTG - Effect size	DRTG - p-value	USG% - Effect size	USG% - p-value
Antetokounmpo	-3.69	0.5229	1.6951	0.6016	-0.652	0.7219
Butler	-4.53	0.3571	-3.783	0.1877	0.7578	0.5656
George	13.604	0.055	1.8056	0.5601	0.6638	0.7149
Mitchell	14.284	0.0706	3.7683	0.7576	0.6397	0.7287
Murray	-1.921	0.7891	-4.255	0.179	-1.249	0.3347
Siakam	4.548	0.4128	-0.516	0.8499	-1.019	0.419

Table 1: Effect size estimates and p-values for Ortg, Drtg, and Usg% for Home factor for our six players

	ORTG - Effect size	ORTG - P-value	DRTG - Effect size	DRTG - p-value	USG% - Effect size	USG% - p-value
Antetokounmpo	5.5902	0.334	8.0529	0.0151*	-0.265	0.8848
Butler	-1.544	0.753	-0.606	0.8319	-0.73	0.5799
George	-2.246	0.7476	4.3056	0.1676	-1.951	0.2851
Mitchell	16.404	0.0385*	6.5839	0.0849	2.0812	0.2611
Murray	16.546	0.0235*	7.895	0.014*	1.4725	0.2561
Siakam	-13.14	0.0201*	-0.743	0.7854	-4.188	0.0014**

Table 2: Effect size estimates and p-values for Ortg, Drtg, and Usg% for Bubble factor for our six players

	ORTG - Effect size	ORTG - P-value	DRTG - Effect size	DRTG - p-value	USG% - Effect size	USG% - p-value
Antetokounmpo	-9.738	0.0947	1.7692	0.5859	-3.793	0.0414*
Butler	-5.085	0.3016	0.1802	0.9497	0.0948	0.9426
George	-4.382	0.5305	-1.094	0.7237	0.0706	0.969
Mitchell	-13.14	0.0955	-1.905	0.6147	0.4849	0.7926
Murray	-3.367	0.6393	0.6716	0.831	-1.115	0.3889
Siakam	-7.325	0.1889	0.7386	0.7866	-1.659	0.1901

*Table 3: Effect size estimates and p-values for Ortg, Drtg, and Usg% for Home*Bubble interaction for our six players*